AD-A096 160   PUBLIC RESEARCH INST   ALEXANDRIA VA   F/G 12/1
ON THE ESTIMATION OF TECHNICAL INEFFICIENCY IN THE STOCHASTIC F--ETC(U)
JAN 81   P SCHMIDT, J JONDROW
UNCLASSIFIED   PRI-PP-300   NL

END
DATE
FILMED
4-81
DTIC

PROFESSIONAL PAPER 300 / January 1981

LEVEL

AD A096160

ON THE ESTIMATION OF TECHNICAL INEFFICIENCY IN THE STOCHASTIC FRONTIER PRODUCTION FUNCTION MODEL

Peter Schmidt
James Jondrow

CNA
The Public Research Institute
A Division of the Center for Naval Analyses

(9) PROFESSIONAL PAPER 300 / Jan[redacted]81

(11)

(2)

(14) PRI-PP-300

(6) # ON THE ESTIMATION OF TECHNICAL INEFFICIENCY IN THE STOCHASTIC FRONTIER PRODUCTION FUNCTION MODEL.

(10) Peter Schmidt*
James Jondrow

*Michigan State University

(12) 13

The Public Research Institute
A Division of the Center for Naval Analyses
2000 North Beauregard Street, Alexandria, Virginia 22311

mt

INTRODUCTION

Consider a production function $y_i = g(x_i, \beta) + \epsilon_i$, $i=1,2,\ldots,N$, where y_i = output for observation i, x_i = vector of inputs for observation i, β = vector of parameters, ϵ_i error term for observation i. The "stochastic frontier" (also called "composed error") model, introduced by Aigner, Lovell, and Schmidt (1977) and Meeusen and van den Broeck (1977), postulates that the error term ϵ_i is made up of two independent components:

$$\epsilon_i = v_i - u_i \tag{1}$$

where $v_i \sim N(0, \sigma_v^2)$ is a two-sided error term representing the usual statistical noise found in any relationship, and $u_i \geq 0$ is a one-sided error term representing technical inefficiency. Note tht u_i measures technical inefficiency in the sense that it measures the shortfall of output (y_i) from its maximal possible value ($g(x_i, \beta) + v_i$).

When a model of this form is estimated, one readily obtains residuals $\hat{\epsilon}_i = y_i - g(x_i, \hat{\beta})$, which can be regarded as estimates of error terms ϵ_i. However, the problem of decomposing such an estimate into separate estimates of the components v_i and u_i has remained unsolved for some time. Of course, the <u>average</u> technical inefficiency--the mean of the distribution of the u_i--is easily calculated. For example, in the half-normal case (u_i distributed as the absolute value of a $N(0,\sigma_u^2)$ variable). The mean technical inefficiency is $\sigma_u\sqrt{2/\pi}$, and this can be evaluated given one's estimate of σ_u, as in Aigner, Lovell, and Schmidt (1977) or Schmidt and Lovell (1979). Or average technical inefficiency can be estimated by the average of the $\hat{\epsilon}_i$. But it is also clearly desirable to be able to estimate the technical inefficiency u_i <u>for each observation</u>.

Intuitively, this should be possible because $\epsilon_i = v_i - u_i$ can be estimated and obviously contains information on u_i. In this paper, we proceed by considering the conditional distribution of u_i given ϵ_i. This distribution contains whatever information ϵ_i yields about u_i. Our estimate of u_i is simply the mean of this conditional distribution, evaluated at $\epsilon_i = \hat{\epsilon}_i$. The variance of the conditional distribution is also calculated, since it indicates how close realizations of the conditional distribution are likely to be to the conditional mean. The conditional means and variances are calculated for the commonly assumed cases of half-normal and exponential u_i, resulting in easily evaluated expressions.

THE HALF-NORMAL CASE

We consider the two-part disturbance given in (1) above, with $v_i \sim N(0,\sigma_v^2)$ and $u_i \sim |N(0,\sigma_u^2)|$. For notational simplicity, we will drop the observation subscript (i) throughout this section.

We define $\sigma^2 = \sigma_u^2 + \sigma_v^2$, $\mu_* = -\sigma_u^2\epsilon/\sigma^2$, $\sigma_*^2 = \sigma_u^2\sigma_v^2/\sigma^2$ Then our main result (proved in the appendix) is the following:

> THEOREM 1: The conditional distribution of u given ϵ is that of a $N(\mu_*,\sigma_*^2)$ variable truncated at zero.

To estimate u, given an estimate $\hat{\epsilon}$ of ϵ, it is natural to pick the mean of the conditional distribution, evaluated at $\hat{\epsilon}$. The mean of a $N(\mu_*,\sigma_*^2)$ variable truncated at zero is (Johnson and Kotz (1970), pp. 81-83):

$$E(u|\epsilon) = \mu_*[1-F(-\mu_*/\sigma_*)] + \sigma_* f(-\mu_*/\sigma_*), \tag{2}$$

where f and F represent the standard normal density and cdf, respectively. But note that

$$-\mu_*/\sigma_* = \epsilon\lambda/\sigma, \tag{3}$$

with $\lambda = \sigma_u/\sigma_v$; this is the same point at which f and F are evaluated in calculating the likelihood function and its derivatives (Aigner, Lovell, and Schmidt (1977, pp. 26-27)). Thus, with simple algebra, we obtain

$$E(u|\epsilon) = \sigma_* \left[\frac{f(\epsilon\lambda/\sigma)}{1-F(\epsilon\lambda/\sigma)} - \left(\frac{\epsilon\lambda}{\sigma}\right)\right]. \tag{4}$$

Replacing ϵ by $\hat{\epsilon}$ gives the desired estimate of u. Incidentally, it is easily verified that the expression in (4) is non-negative, and monotonic in ϵ.

The estimate of u obtained by evaluating (4) at $\hat{\epsilon}$ contains two types of variance. The first is ordinary sampling error, due to the variability of $\hat{\epsilon}$ as an estimate of ϵ (which in turn is due to the variability of $\hat{\beta}$ as an estimate of β). This variability disappears asymptotically and can, therefore, be ignored for large enough sample sizes. However, the estimate in (4) still contains the variance of the conditional distribution, which is independent of sample size and should not be ignored; it is a reflection of the obvious fact that ϵ contains only imperfect information about u. The conditional variance is simply the variance of a $N(\mu_*,\sigma_*^2)$ variable truncated at zero, which is

$$\mathrm{Var}(u|\epsilon) = \sigma_*^2\left\{1 + \frac{\epsilon\lambda}{\sigma}\,\frac{f(\epsilon\lambda/\sigma)}{1-/f(\epsilon\lambda/\sigma)} - \left[\frac{f(\epsilon\lambda/\sigma)}{1-F(\epsilon\lambda/\sigma)}\right]^2\right\} \quad (5)$$

Thus, (5) gives a measure of the variability of the estimate (4), for sample sizes large enough to ignore the sampling error. Of course, since we are dealing with a truncated normal distribution, we should be careful in our interpretation of standard errors;

confidence intervals for u might better come from the conditional distribution (Theorem 1) itself than from consideration of u relative to its standard error.

THE EXPONENTIAL CASE

This case is identical to the half-normal case, except that now the technical inefficiency error term u is assumed to follow the one-parameter exponential distribution with density

$$f(u) = \exp(-u/\sigma_u)/\sigma_u \qquad (6)$$

Our results are similar to those for the half-normal case. Define $A = \epsilon/\sigma_v + \sigma_v/\sigma_u$. Then we have the following result, proved in the appendix:

> THEOREM 2: The conditional distribution of u given ϵ is that of a $N(-\sigma_v A, \sigma_v^2)$ variable, truncated at zero.

The mean and variance of this distributon are therefore exactly as given in (4) and (5) if we just replace σ_* by σ_v and replace $\epsilon\lambda/\sigma$ by A. As in the half-normal case, these would be evaluated at $\epsilon = \hat{\epsilon}$.

AN EXAMPLE

Schmidt and Lovell (1980) estimated a system consisting of a stochastic frontier production function and first-order conditions for cost minimization, based on a sample of 111 steam-electric generating plants. The estimates on which our calculations are based are those reported in the first column of table 1 of Schmidt and Lovell. In particular, note that $\hat{\sigma}_u^2$ = .014452, $\hat{\sigma}_v^2$ = .003261, and that the estimated average technical inefficiency (mean of u) is .09592, indicating about 9.6 percent technical inefficiency.

We have calculated $\hat{u} = E(u|\hat{\epsilon})$ for each observation, based on (4) since estimation assumed half-normal u, as well as the variance of this estimate based on (5). We will not present results for all 111 observations, but rather, point out a few things of interest. The mean of the u's is .08387, which is in the same ballpark as the .09592 reported above and as the mean of -.07433 for the $\hat{\epsilon}$'s. The smallest $\hat{u}$ (most technically efficient plant) was .01658, with a standard error of .01534, based on $\hat{\epsilon}$ = .15891. This is a modest outlier, with $\hat{\epsilon}$ about 2.75 standard deviations from the mean $\hat{\epsilon}$. (The other observations had $\hat{u}$ under .02 also.) The most technically efficient observations can be characterized as having high outputs (the two most efficient plants have the second and third largest outputs in the sample), low capital stocks and high levels of fuel consumption and labor usage. Their level of allocative inefficiency (see Schmidt and Lovell (1979)) is below average, though not strongly so. On the other hand, the largest $\hat{u}$ (most technically inefficient plant) was .37156, with a standard error of .05105, based on $\hat{\epsilon} = -.45540$. This is a fairly large outlier, in the sense that $\hat{\epsilon}$ is almost 4 standard deviations from the mean $\hat{\epsilon}$; but one other observation had an almost identical value of $\hat{u}$ and $\hat{\epsilon}$. The most technically inefficient observations have rather average outputs but higher than average uses of capital, fuel, and labor. They also had slightly above average levels of allocative inefficiency.

CONCLUSIONS

In this paper, ~~we have proposed~~ a method is proposed for ~~of~~ separating the error term of the stochastic frontier model into its two components for each observation. This method enables one to estimate the level of technical inefficiency for each observation in the sample, and largely removes what had been viewed as a considerable disadvantage of the stochastic frontier model relative to other models (so-called deterministic frontiers) for which technical inefficiency was readily measured for each observation.

APPENDIX

THE HALF-NORMAL CASE

In the half-normal case, $v \sim N(0,\sigma_v^2)$, u is distributed as the absolute value of $N(0,\sigma_u^2)$, v and u are independent, and $\epsilon = v-u$. We wish to find the distribution of u conditional on ϵ.

The joint density of u and v is the product of their individual densities, since they are independent:

$$f(u,v) = \frac{1}{\pi\sigma_u\sigma_v} \exp\left[\frac{-1}{2\sigma_u^2} u^2 - \frac{1}{2\sigma_v^2} v^2\right], \quad u \geq 0 . \qquad (A1)$$

Making the transformation $\epsilon = v-u$, the joint density of u and ϵ is

$$f(u,\epsilon) = \frac{1}{\pi\sigma_u\sigma_v} \exp\left[- \frac{1}{2\sigma_u^2} u^2 - \frac{1}{2\sigma_v^2} (u^2+\epsilon^2+2u\epsilon)\right]. \qquad (A2)$$

The density of ϵ is given by equation (8) of Aigner, Lovell, and Schmidt (1977):

$$f(\epsilon) = \frac{2}{\sqrt{2\pi}\,\sigma} (1-F) \exp\left[- \frac{1}{2\sigma^2} \epsilon^2\right] \qquad (A3)$$

where $\sigma^2 = \sigma_u^2 + \sigma_v^2$, $\lambda = \sigma_u/\sigma_v$, and F is the standard normal cdf, evaluated at $\epsilon\lambda/\sigma$. Therefore, the conditional density of u given ϵ is the ratio of (A2) to (A3), which we can write as

$$f(u\,|\epsilon) = \frac{1}{\sqrt{2\pi}\,\sigma_*}\,\frac{1}{1-F}\,\exp\left[\frac{-1}{2\sigma_*^2}\,u^2 - \frac{1}{\sigma_v^2}\,u\epsilon - \frac{\lambda}{2\sigma^2}\,\epsilon^2\right],\ u\geq 0, \quad (A4)$$

where $\sigma_*^2 = \sigma_u^2\sigma_v^2/\sigma^2$. With a little algebra, this simplifies to

$$f(u\,|\epsilon) = \frac{1}{1-F}\,\frac{1}{\sqrt{2\pi}\sigma_*}\,\exp\left[\frac{-1}{2\sigma_*^2}\,(u+\sigma_u^2\epsilon/\sigma^2)^2\right],\ u\geq 0. \quad (A5)$$

Except for the term involving 1-F, this looks like the density of $N(\mu_*,\sigma_*^2)$, with $\mu_* = -\sigma_u^2\epsilon/\sigma_*$. Finally, note that F is evaluated at $\epsilon\lambda/\sigma = -\mu_*/\sigma_*$, and thus (1-F) is just the probability that a $N(\mu_*,\sigma_*^2)$ variable be positive. Thus, (A5) is indeed the density of a $N(\mu_*,\sigma_*^2)$ variable truncated at zero.

The Exponential Case

In the exponential case, $v \sim N(0,\sigma_u^2)$ while u is independent of v and has the one-parameter exponential density

$$f(u) = \frac{1}{\sigma u}\exp(-u/\sigma_u), \quad u\geq 0 \quad (A6)$$

with $\epsilon = v-u$, the joint density of u and ϵ, becomes

$$f(u,\epsilon) = \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \exp\left[- \frac{1}{2\sigma_v^2} u^2 - \left(\frac{\epsilon}{\sigma_v^2} + \frac{1}{\sigma_u} \right) u\epsilon - \frac{1}{2\sigma_v^2} \epsilon^2 \right] \quad (A7)$$

From Aigner, Lovell, and Schmidt (1977, p. 29), the density of ϵ is

$$f(\epsilon) = \frac{1-F}{u} \exp\left[\frac{\epsilon}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2} \right] \quad (A8)$$

where F is again the standard normal cdf, but evaluated at

$$A = \epsilon/\sigma_v + \sigma_v/\sigma_u \quad (A9)$$

Taking the ratio of (A7) to (A8) and simplifying,

$$f(u|\epsilon) = \frac{1}{1-F} \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left[- \frac{1}{2\sigma_v^2} \left(u + \sigma_v A \right) \right], \ u>0. \quad (A10)$$

But this is just the density of a $N(-\sigma_v A, \sigma_v^2)$ variable, truncated at zero.

REFERENCES

[1] Aigner, D.J., C.A.K. Lovell, and P. Schmidt (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models," Journal of Econometrics 6, 21-37.

[2] Johnson, N.L. and Skotz (2970), Continuous Univariate Distributions -1, Boston: Houghton Miffin Company

[3] Meeusen, W. nd J. van den Broeck (1977), "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error," International Economic Review 18, 435-444.

[4] Schmidt, P. and C.A.K. Lovell (1979), "Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and Cost Frontiers," Journal of Econometrics 9, 343-366.

[5] Schmidt, P. and C.A.K. Lovell (1980), "Estimating Stochastic Production and Cost Frontiers when Technical and Allocative Inefficiency are Correlated," Journal of Econometrics 13, 83-100.

FILMED

4-8